



A ROADMAP TO DETERMINE THE IMPORTANT FACTORS OF THE HOUSE VALUE: A CASE STUDY BY USING ACTUAL PRICE REGISTRATION DATA OF TAIPEI HOUSING TRANSACTIONS

Mingchin Chen
Fu Jen Catholic University, Taiwan
E-mail: 081438@mail.fju.edu.tw

Pei-De Wang
Fu Jen Catholic University, Taiwan
E-mail: kmpeterwang@gmail.com

Submission: 08/06/2017
Revision: 29/06/2017
Accept: 22/07/2017

ABSTRACT

While many studies have applied data mining techniques to judge housing prices, few have decoded the important attributes or prioritized them simultaneously. This paper aims to utilize five data mining techniques to discover the important attributes for three major types of real estate in Taipei city. The datasets, involving a total of 22,480 transactions, were publicly available from the Taiwan Actual Price Registration from July 2013 to August 2015. The five models are decision trees, random forests, model trees, artificial neural networks and multiple regression. The criteria used to measure the forecasting accuracy are MAPE, R^2 , RMSE, MAE and COR. The model with the best performance for all houses is the Model Tree with a MAPE value of 27.59. As for apartments, the best is Random Forests. Artificial Neural Networks perform best for suites and buildings with elevators. Different housing types need different models. Furthermore, the attributes importance helps us to conclude the really critical attributes, which include the floor area, administrative districts, parking area and land area, and their rankings. This variable ranking and selection procedure proposed by this research can also be adopted to improve the prediction efficiency for most big data applications other than the housing transactions.



Keywords: data mining; housing pricing; forecasting accuracy; variables ranking; variables selection

1. INTRODUCTION

Buying a house in Taipei is relatively hard-affordable. Therefore, evaluating a housing price become an issue. Even Taiwan authorities take the transactions more transparent in action. Taipei remains one of the most expensive cities in the world in which to buy a house. Taipei's house price-to-income ratio stood between 15 and 17 in 2015, higher than London (8.5x), New York (5.9x), or Sydney (12.2x) (DELMENDO, 2016).

Housing affordability remains a major problem in Taipei city. Furthermore, higher housing affordability means higher housing prices relatively. In addition, there must be some inherent factors giving rise to these high housing prices. Those inherent factors determine the housing prices and meanwhile stand for the favor of people when they are going to buy or sell a house in Taipei.

Actual Price Registration (APR) refers to a national system for registering the actual prices of property transactions—an initiative created to boost transparency in Taiwan's real estate market. This regulation came into effect on August 1, 2011. This study intends to determine what those factors are from that open system with real transactions by utilizing five data mining skills.

There are 3 major housing types to which this paper pays particular attention. According to statistics from the Department of Urban Development, Taipei City Government for 2013 to 2015, transactions involving condominiums in buildings of 5 storeys or less without an elevator (apartment) accounted on average for 21% of housing transactions (Type_APT), condominiums with elevators (buildings) for 58% (Type_BLD) and suites (Type_SUT) for 19% as shown in Figure 1.

The curve corresponding to the right coordinate axis represents the volumes of transactions in each season. Even though the volumes have changed over time, the percentages of those 3 types remain relatively stable. Therefore, those 3 types become our study targets.

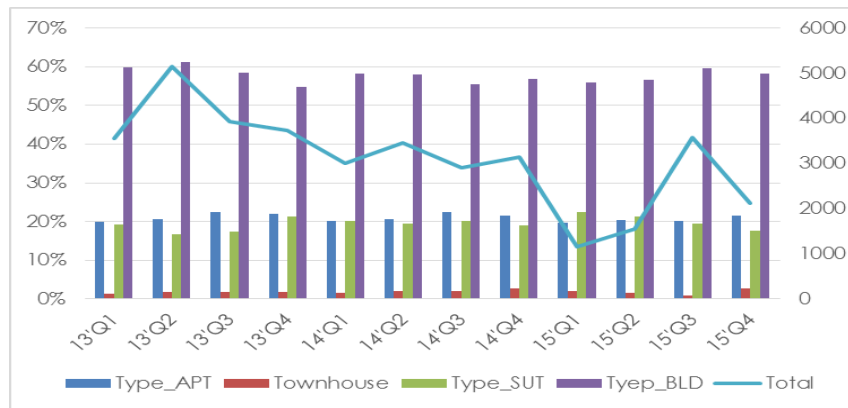


Figure 1: Volumes of transactions from 2013 to 2015

The hedonic-based regression approach has been utilized extensively to investigate the relationship between house prices and housing characteristics (FAN; ONG; KOH, 2006). For example, Goodman (1978) extended hedonic price analysis to the formation of housing price indices measuring variations within a metropolitan area (GOODMAN, 1978).

Fit et al., (2003) developed several hedonic specifications that attempt to more fully capture the interactive components of location values (FIK; LING; MULLIGAN, 2003). Welch et al., (2016) estimated a hedonic spatial panel model to determine the long-term impact of improved network access to bike and public transit facilities on housing sales prices (WELCH; GEHRKE; WANG, 2016). However, this approach is subject to criticisms arising from potential problems related to fundamental model assumptions and estimation (FAN; ONG; KOH, 2006).

Nowadays, there are more and more studies that focus on real estate by using data mining techniques. Acciani et al., (2011) adopted model trees and multivariate adaptive regression splines to predictors in real estate appraisal (ACCIANI; FUCILLI; SARDARO, 2011). Fong and Wah (2013) utilized feature selection techniques to screen important attributes and applied those attributes to build up a predictive model by using different kinds of data mining techniques. Gan et al., (2015) built decision trees and neural networks and compared their results.

While these authors all used different data mining techniques to figure out the housing prices, few of them attempted to find out what were the important attributes or to rank them by importance at the same time. Moreover, none of them identified the attributes according to the types of houses.

This paper is going to utilize five models and five measurements to evaluate them. The five models are decision trees, random forests, model trees, artificial neural networks and multiple regression. The criteria used to measure the forecasting accuracy are MAPE, R^2 , RMSE, MAE and COR. The final result is the roadmap for evaluating the more reasonable housing prices.

2. RESEARCH METHODOLOGY

The research flow is shown in figure 2. All the data used in this paper is downloaded from APR. By using 5 data mining techniques and comparing 3 major housing types by MAPE, R^2 , RMSE, MAE and COR. This paper finds out that different housing types need different data mining models.

Each type has its own favor attributes with higher importance values. Therefore, ranks those attributes according to the averages of these importance values. Then count the number of models that have the same attributes. This ranking and selection process helps us to figure out the relative important attributes in each housing types.

Finally, according to the statistics on rankings and votes of attributes, this paper identify the classifications of the attributes and build a roadmap to depict the diversities of attributes.

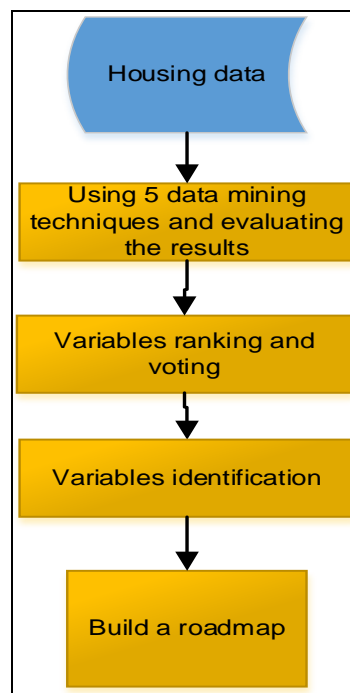


Figure 2: Research flow

3. DATA MINING SKILLS

This session is going to introduce the data mining skills used in this paper.

3.1. Decision Tree(DT)

A DT algorithm works by splitting a dataset in order to build a model that successfully classifies each record in terms of a target field or variable (WOODS; KYRAL, 1997). There are two types of DT: a classification tree and a regression tree that can be implemented using the four most popular algorithms: the chi-squared automatic interaction detection (CHAID) (KASS, 1980; MAGIDSON, 1994), the iterative dichotomiser (ID3) (QUINLAN, 1986), the classification and regression trees (BREIMAN; FRIEDMAN; OLSHEN; STONE, 1984) and C4.5 (QUINLAN, 1992).

CHAID and ID3 can only be used for the classification tree, while both the classification and regression trees can be used for the others. A response variable which has more classes or categories than a classification tree can be used, otherwise a regression tree that has numeric or continuous responsiveness can be used instead.

Two main processes used to construct a tree are tree growing and pruning. The tree growing process searches for independent variables as splitters that start from the root node with all the instances and keeps partitioning those with the greatest differences until no significant differences can be identified. In this process, the purity or impurity criterion is used to split a node that makes instances more likely in a node. In the case of a classification tree, splitting the data is based on homogeneity. A regression model splits each of the independent variables as nodes where their inclusion decreases the error measure the most. The best criterion should produce the greatest purity or reduce the impurity the most.

3.2. Random Forests (RF)

The pros and cons of DT are as follows (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2013). The advantages are that they are easy to explain, more closely mirror human decision-making, may be displayed graphically and can easily handle qualitative predictors. Unfortunately, DT generally do not have precise predictive power. However, the performance of the predictive power can be substantially improved by RF.

In actual fact, RF are an example of ensemble methods that combines a series of k base models (or trees) with the aim of co-creating an improved composite model. Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (BREIMAN, 2001). After a large number of trees are generated, they are combined to yield a single consensus prediction by voting for classification trees or averaging for regression trees. Besides, RF are characterized by significant improvements in accuracy, and greater robustness to errors and outliers.

There are two basic beliefs regarding RF in that most trees can provide correct predictions and the trees make mistakes in different places. Beriman (2001) stated that the use of the Strong Law of Large Numbers shows that RF always converge so that overfitting is not a problem and they produce limit values of the generalization errors that are measures of how accurate the individual classifiers are (strength) and of the dependence between them (correlation) (BREIMAN, 2001). The idea is to maintain the strength without increasing their correlation.

3.3. Model Tree (MT)

The MT is based on a divide-and-conquer approach through which it is possible to learn from a set of instances (WITTEN; FRANK, 2005). The output of a MT is represented by a tree-like structure in which it is possible to distinguish a root node, parent and child nodes, arches (or branches) and leaves (ACCIANI; FUCILLI; SARDARO, 2011).

The greatest difference when compared with a decision tree is the content of the leaf node. In the model tree, each terminal node represents more and delivers more information. A linear regression model is calculated based on the number of instances of that node that it contains, and not on an averaged value in the regression tree. As a result, it may provide a more precise estimation. This paper uses a rule-based model that is an extension of Quinlan's M5 MT (KUHN; WESTON; DEEFER; COUTLER, 2016).

3.4. Artificial Neural Network (ANN)

ANN is an artificial intelligence model originally designated to replicate the human nervous system (BAHIA, 2013). Once the nervous system is alerted by outside stimulations, neurons work and react. Therefore, ANN consists of three main

layers: the input data layer (stimulations), the hidden layer(s) and the output layer. Each artificial neuron has a set of input connections that receive signals from other neurons and a bias adjustment, as well as a transfer function that transforms the sum of the weighted inputs and bias to decide the value of the output (COAKLEY; BROWN, 2000).

3.5. Multiple Regression (MR)

The hedonic-based regression approach belongs to MR. There are many independent variables and one dependent variable in MR. The relationships between the independent variables and dependent variable will be described. Fixed independent variables derive the conditional expectation of the dependent variable, an averaged value. Therefore, MR is widely used for prediction.

4. DATA SOURCE AND PREPARATION

The data used in this research are downloaded from APR. Raw data amount to 48,658 observations from July 2013 to August 2015. After deleting all records with empty column(s) and unreasonable values, the total number of observations is 22,480 and encompasses the three most popular housing types that are all only for home use.

To facilitate further inspections and comparisons, this paper also combines each of these three types into an overall group (Type_ALL). Generally speaking, Type_APT and Type_BLD are both suitable for a family and Type_SUT might be more suitable for singles.

There are 20 attributes used in this paper that are listed in Table 1. This research has partitioned the houses into three types, and therefore the total number of attributes used in Type_APT, Type_BLD and Type_SUT is 19. The housing prices are naturally chosen as the dependent variable while the other housing attributes are treated as independent variables.

There are two types of attributes: C stands for category and N for numeric. The amounts of data used in Type_APT, Type_BLD and Type_SUT are 6,115, 13,039 and 3,326, respectively. Two-thirds of the sample data are used in building the model, and the remaining one-third is used as an external holdout for measurement purposes.



Table 1: Data attributes

Attributes	Type	Description
1 target_dst	C	Administrative districts: Songshan(1), Sinyi(2), Da-an(3), Jhongshan(4), Jhonjheng(5), Datong(6), Wanhua(7), Wunshan(8), Nangang(9), Neihu(10), Shihlin(11) and Beitou(12).
2 target_tp	C	With(1) or without(2) parking place
3 Ind_area	N	Occupied land area of the house(M ²)
4 Indusg_tp	C	Type of land usage: Residential(1), Commercial(2), Industrial(3), Others(4), Agricultural(5)
5 ym_sold	C	Year and month when the house has been sold
6 prk_sold	N	Number of parking places sold
7 flat_type	C	Floor numbering
8 total_flat	N	Total floor level of a building
9 hs_tp	C	Housing types: APT(1),BLD (2) and SUT(6)
10 cstrct_tp	C	Types of construction methods: Reinforced concrete (1),Reinforced brick structure (2), Referring to building occupation permit (3), Brick structure (4), Steel reinforced concrete (5), Referring to other registrations (6), Steel concrete (7), Precast reinforced concrete (8).
11 flr_area	N	Area of the house (M ²)
12 room	N	Number of rooms
13 sit_room	N	Number of living and/or dining rooms
14 bathroom	N	Number of bathrooms
15 cmptmt	C	Compartment (1) or not (2)
16 mgt_cmt	C	Having (1) or not having (2) a management committee
17 pk_type	C	Parking type: No parking space (0), On the ground floor (1), Lifting plane (2), Lifting machinery (3), Ramp (4), Ramp machinery (5), Tower (6), Others (7)
18 pk_area	N	Parking area (M ²)
19 flat_age	N	Housing age (year)
20 price	N	Total price (NTD)

5. RESULTS AND DISCUSSION

The purpose of this section is to ascertain the predominant attributes of housing prices. Five models are utilized in the prediction. There are many criteria used to measure the forecasting accuracy (MUNUSAMY; MUTHUVEERAPPAN; BABA; ABDULLAH; ASMONI, 2015).



In this paper, the measures used for comparison purposes are the MAPE (Mean Absolute Percentage Error), R² (Coefficient of determination), RMSE (Root mean squared error), MAE (Mean absolute error) and COR (Correlation).

The results are derived from package 'rminer' (CORTEZ, 2016) and displayed in Table 2. The notation "<" means "better" if a lower value, and ">" stands for "better" if a higher value. The notation "1" represents the best performance based on the specific measure for each housing type.

For all houses, the MT's MAE is larger than the RF's, however, the MT's RMSE is smaller than the RF's. That means RF have more forecasting values closer to real prices than the MT, but meanwhile the RF have more outliers than the MT. In a word, MT has the best forecasting performance of all houses because the MT has the four best measures of the five.

For apartments, RF have all the measures to win: the smallest MAPE, RMSE and MAE, and the largest R² and COR. Furthermore, ANN is found to do better than the other models because over half the measures are better than those for the other models. Obviously, due to the distinct characteristics of the different housing types, different algorithms need to be adopted.

Table 2: Measurement results for all types

Model	Measurement	Type_ALL	Type_APT	Type_BLD	Type_SUT
DT	MAPE "<"	54.73	62.11	48.75	31.78
ANN		34.38	46.83	32.00	20.81
MR		50.58	50.18	40.93	23.94
MT		27.59 ¹	41.23	27.97 ¹	16.90 ¹
RF		27.95	39.71 ¹	29.33	23.38
DT	R ² ">"	0.73	0.49	0.68	0.61
ANN		0.81	0.53	0.87 ¹	0.80 ¹
MR		0.73	0.57	0.74	0.76
MT		0.84 ¹	0.56	0.84	0.78
RF		0.78	0.59 ¹	0.80	0.79
DT	RMSE "<"	12447320	5003089	15617442	2789666
ANN		10374326	4794231	9971407 ¹	1996698 ¹
MR		12349125	4565800	14159109	2196543
MT		9648036 ¹	4628516	11232151	2010710

Model	Measurement	Type_ALL	Type_APT	Type_BLD	Type_SUT
RF		11181599	4479439¹	12480842	2032913
DT	MAE "<"	6638674	3642247	8678978	1986994
ANN		4531948	3436007	5762095	1335743 ¹
MR		6051855	3321844	7791333	1536106
MT		4307236	3289040	5485538	1372310
RF		4115843¹	3169619¹	5467660¹	1384466
DT		COR ">"	0.85	0.70	0.83
ANN	0.90		0.74	0.93¹	0.89
MR	0.86		0.76	0.86	0.87
MT	0.92¹		0.75	0.92	0.89
RF	0.90		0.78¹	0.91	0.90¹

Each type has its own ranking or focused attributes. Insights may be gained by utilizing the important values of each attribute in a model that can be derived from rminer. Different models have different importance values for the same attributes. Inspired by the ensemble model, this paper averages those importance values from the five models outlined above, and ranks those attributes according to the averages of these importance values.

Those attributes appearing in the bold frames constitute 95 percent of the importance resident in each type as shown in Table 3. We then count the number of models (#) that have the same attributes among the top 10 attributes of Type_ALL and these appear in the bold frames for each method simultaneously. The results can be seen as the voting results based on the five models.

For Type_BLD, for instance, there are nine attributes that account for 95 percent of the importance with respect to housing prices. Those attributes from the most important to the least important are floor area, land area, number of rooms, and number of sold parking places, etc. The attribute of floor area in Type_BLD receives the five models' votes, land area gets four, and number of rooms gets two, and so on.

Table 3: Rankings and votes of attributes

Ranking	Type_ALL		Type_APT		Type_BLD		Type_SUT	
	Attributes	#	Attributes	#	Attributes	#	Attributes	#
1	flr_area	5	flr_area	5	flr_area	5	flr_area	5
2	target_dst	4	target_dst	5	lnd_area	4	target_dst	5
3	pk_area	3	bathroom	4	room	2	flat_age	5
4	lnd_area	2	lnd_area	4	prk_sold	3	lnd_area	4
5	room	2	pk_area	3	target_dst	5	cstrct_tp	3
6	prk_sold	3	sit_room	4	pk_area	3	pk_type	4
7	total_flat	3	prk_sold	2	total_flat	2	total_flat	2
8	bathroom	1	flat_age	3	bathroom	3	pk_area	4
9	cstrct_tp	3	target_tp	2	cstrct_tp	2	prk_sold	1
10	cmptmt	1	room	4	flat_age	2	Indusg_tp	2
11	sit_room	2	cstrct_tp	2	pk_type	1	bathroom	3
12	flat_age	2	ym_sold	1	sit_room	1	room	3
13	target_tp	3	cmptmt	3	target_tp	1	cmptmt	2
14	pk_type	1	pk_type	2	cmptmt	0	flat_type	2
15	Indusg_tp	2	Indusg_tp	2	flat_type	0	target_tp	1
16	ym_sold	1	flat_type	1	Indusg_tp	0	sit_room	0
17	flat_type	1	mgt_cmt	1	mgt_cmt	0	ym_sold	0
18	hs_tp	0	total_flat	0	ym_sold	0	mgt_cmt	0
19	mgt_cmt	0						

Floor area is the most important and most robust attribute, and all five models agree with the three types. There are many studies whose findings are in line with this point of view. Sirmans et al. (2006) stated that floor area is perhaps the most important structural attribute in determining house prices (SIRMANS; MACDONALD; MACPHERSON; ZIETZ, 2006). In addition, Bracke (2015) showed that the contribution of floor area is positive for housing prices. Xiao et al. (2016) also said that property prices increase as floor area increases.

Moreover, to discover the characteristics of these attributes, this paper extracts the 10 most important attributes from Type_ALL in Table 3 and uses these attributes as the baseline. For each type of house, we sum each model's votes (#), sum the rankings of each attribute before averaging them, and, finally, calculate the variances of the rankings as shown in Table 4.

Table 4: Statistics on rankings and votes

Ranking	Attributes	Sum of Votes	Sum of Rankings	Averaged Rankings	Variances of Rankings
1	flr_area	15	3	1.00	0.00
2	target_dst	15	9	3.00	4.50
3	pk_area	10	19	6.33	0.50
4	Ind_area	12	10	3.33	2.00
5	room	9	25	8.33	24.50
6	prk_sold	6	20	6.67	4.50
7	total_flat	4	32	10.67	60.50
8	bathroom	10	22	7.33	12.50
9	cstrct_tp	7	25	8.33	2.00
10	cmtmt	5	40	13.33	0.50

There are various other inferences obtained from these attributes. By identifying the attributes, these inferences will be discovered. Those attributes in Table 4 that occupy over 50 percent of total votes (15) are referred to as major. Meanwhile, those that have relatively small variances of rankings (less than 5) are referred to as stable.

Thus the major-stable attributes are identified in red shading, such as floor area, administrative districts, parking and land area, due to their high importance and relatively small variances. Similarly, those major-unstable attributes appear with orange shading, the minor-stable ones with yellow and the minor-unstable ones with green.

First, del Cacho (2010) stated that location is a factor of paramount importance when determining the pricing of a property. Second, in downtown areas and inner cities, parking requirements could profoundly alter the housing stock (MANVILLE, 2013). Therefore, parking requirements can increase the price of real estate (SHOUP, 2014). Finally, a larger land area leads to more floor area in each of those three types of housing. Therefore, land area is also an indicator.

Furthermore, the attributes that are referred to as type-dependent attributes show up in the bold frames for Type_APT and Type_SUT, but do not appear in the bold frames for Type_ALL in Table 3. This indicates that different types have their own favorite attributes in addition. Finally, there are attributes outside the bold area for each type of housing that are referred to as others. Those attributes are less important.

By identifying the attributes, the roadmap of importance as shown in Figure 3 is constructed. This could serve as a reference when people appraise a house in Taipei. For example, when people want to buy a condominium with an elevator, the first considerations will be floor area, target district, parking area and land area, all of which are major-stable attributes. Next, major-unstable attributes, such as the numbers of rooms and bathrooms, followed by minor attributes, will be taken into account. Finally, other attributes will be considered.

The roadmap depicts the diversities of attributes. The same type of major-unstable attributes, for example, the number of rooms and bathrooms, appears in different ranking positions. The apartments and condominiums with an elevator are preferred in terms of the number of rooms and bathrooms than the suites. This road map helps us to price the houses.

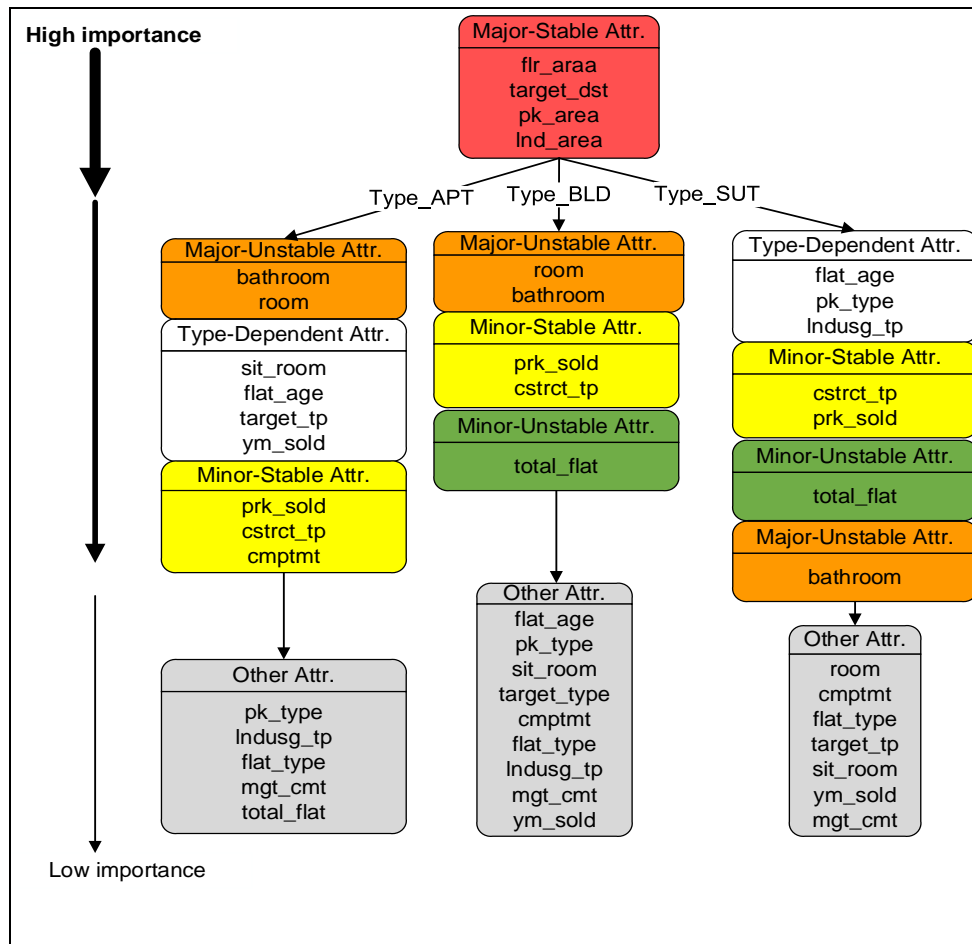


Figure 3: Roadmap of important attributes

The attributes in the bold area may or may not always be important. In view of this, we captured those attributes in the bold frames in Table 3 and reran those five

models. The total amounts of the independent attributes used in Type_APT', Type_BLD' and Type_SUT' were changed to 13, 9 and 12, respectively.

Those attributes were considered to be the most important 95 percent from the appraisals of the five models for each housing type. The consequences are listed in Table 5. The yellow shadings reflect belonging to the better parts of the performances than in the previous experiment that adopted all 19 attributes in the evaluation. The green parts were worse and the white parts were equal.

Type_SUT' performs better in the situation where only 12 important attributes were used. This indicates that most of the important attributes for Type_SUT' were found in this paper. However, those attributes for Type_APT' and Type_BLD' did not work as well as those for Type_SUT'. This reveals that there are attributes that were considered to be more important than this research discovered that were not exposed.

Table 5: Measurement results for 3 major types

Model	Measurement	Type_APT'	Type_BLD'	Type_SUT'
DT	MAPE "<"	62.11	48.57	31.03
ANN		45.76	33.66	16.59
MR		50.18	44.07	23.76
MT		41.52	28.36	16.99
RF		40.05	27.13	16.90
DT	R ² ">"	0.49	0.71	0.66
ANN		0.56	0.83	0.81
MR		0.56	0.73	0.76
MT		0.56	0.83	0.78
RF		0.58	0.83	0.83
DT	RMSE "<"	5003089	15051426	2581766
ANN		4621929	11397272	1918152
MR		4620246	14371934	2196793
MT		4612869	11513048	2090838
RF		4530046	11428551	1840169
DT	MAE "<"	3642247	8492579	189946
ANN		3278997	6381590	1254615
MR		3324592	7992023	1530882

MT		3298646	5671468	1378868
RF		3205514	5095230	1190809
DT	COR ">"	0.70	0.84	0.82
ANN		0.75	0.91	0.90
MR		0.75	0.86	0.87
MT		0.75	0.91	0.89
RF		0.77	0.92	0.91

6. CONCLUSION

In this study, five data mining techniques were constructed from the Actual Price Registration of Taiwan to examine those models' performances in regard to prediction, and to find out the relatively important attributes that will help to identify which attributes are more important according to the type of houses. In such a big data era with huge volumes of data, variables and methods, this paper delineates a road map for the selection of variables in relation to house prices.

First, this paper used five measures, namely, the MAPE, R^2 , RMSE, MAE and COR, to evaluate those five models' performances in terms of prediction. In general, there was no one single best model that could satisfy all three types of houses concurrently. While random forests were more suitable for apartments, ANN were more reliable for the condominiums with elevator(s) and for the suites. The only reason for this was that the patterns of each housing type were not completely similar. Therefore, the model selected was dependent on the housing type.

Second, Figure 3 will help us to identify which attributes are important and their rankings. Through the process of identification, influential factors will be shown in sequence, and decisions to buy or set prices will be made.

Suggestions for future studies include vicinity issues, such as the distances to schools, department stores and parks, etc. That should be taken into account. This research lacks this kind of information. However, the models used could be revalidated when having such data. More new findings about the neighborhood of the houses will be obtained.

REFERENCES



- ACCIANI, C.; FUCILLI, V.; SARDARO, R. (2011) Data Mining in Real Estate Appraisal: A Model Tree and Multivariate Adaptive Regression Spline Approach. **Aestimum**, v. 58, p. 27-45.
- BAHIA, I. S. H. (2013) A Data Mining Model by Using ANN for Predicting Real Estate Market: Comparative Study. *International Journal of Intelligence Science*, v. 3, n. 4. p. 162-169.
- BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. (1984) *Classification and Regression Trees*, Belmont, CA: Wadsworth.
- BREIMAN, L. (2001) Random Forests. *Machine Learning*, v. 45, n. 1, p. 5-32.
- BRACKE, P. (2015) House Prices and Rents: Microevidence From A Matched Data Set in Central London. *Real Estate Economics*, v. 43, n. 2, p. 403-431.
- COAKLEY, J. R.; BROWN, C. E. (2000) Artificial Neural Networks in Accounting and Finance: Modeling Issues. **International Journal of Intelligent Systems in Accounting, Finance and Management**, v. 9, n. 2. p. 119-144.
- CORTEZ, P. (2016) Package 'rminer'. Available: <https://cran.r-project.org/web/packages/rminer/rminer.pdf> . Access: 2th September, 2016.
- DEL CACHO, C. (2010) A Comparison of Data Mining Methods for Mass Real Estate Appraisal (No. 27378). **Munich Personal RePEc Archive**.
- DELMENDO, L. C. 2016. Taiwanese House Prices Continue to Fall Due to Harsh Taxes. Retrieved on September 16, 2016, from <http://www.globalpropertyguide.com/Asia/Taiwan/Price-History>
- FAN, G. Z.; ONG, S. E.; KOH, H. C. (2006) Determinants of House Price: A Decision Tree Approach. **Urban Studies**, v. 43, n. 12, p. 2301-2315.
- FIK, T. J.; LING, D. C.; MULLIGAN, G. F. (2003) Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach. **Real Estate Economics**, v. 31, n. 4, p. 623-646.
- FONG, S.; WAH, Y. B. (2013) A Prediction Model for Forecasting the Trend of Macau Property Price Movements and Understanding the Influential Factors. **Journal of Emerging Technologies in Web Intelligence**, v.5, n. 2, p. 122-131.
- GAN, V.; AGARWAL, V.; KIM, B. (2015) Data Mining Analysis and Predictions of Real Estate Prices. **Issues in Information Systems**, v. 16, n. 4, p. 30-36.
- GOODMAN, A. C. (1978) Hedonic Prices, Price Indices and Housing Markets. **Journal of Urban Economics**, v. 5, n. 4, p. 471-484.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. (2013) **An Introduction to Statistical Learning**, New York: Springer.
- KASS, G. V. (1980) An Exploratory Technique for Investigating Large Quantities of Categorical Data. **Applied Statistics**, v. 29, n. 2, p. 119-127.
- KUHN, M.; WESTON, S.; DEEFER, C.; COUTLER, N. (2016) Cubist Models for Regression, Available: <https://cran.r-project.org/web/packages/Cubist/vignettes/cubist.pdf> . Access: 10th December, 2016.
- MAGIDSON, J. (1994) The CHAID Approach to Segmentation Modeling: Chi-squared Automatic Interaction Detection, in: BAGOZZI, R. P. (Ed.), **Advanced**

Methods of Marketing Research. Malden (Mass. US): Blackwell Business, p. 118-159.

MANVILLE, M. (2013) Parking Requirements and Housing Development: Regulation and Reform in Los Angeles. **Journal of the American Planning Association**, v. 79, n. 1, p. 49-66.

MULLEY C. (Ed.), **Parking: Issues and Policies.** United Kingdom: Emerald Publishing, p. 87-113.

MUNUSAMY, M.; MUTHUVEERAPPAN, C.; BABA, M.; ABDULLAH, M. N.; ASMONI, M. (2015). An Overview of the Forecasting Methods Used in Real Estate Housing Price Modelling. **Jurnal Teknologi**, v. 73, n. 5, p. 189-193.

QUINLAN, J. R. (1986) Induction of Decision Trees. **Machine Learning**, v. 1, p. 81-106.

QUINLAN, J. R. (1992) **C4. 5: Programming for Machine Learning**, San Mateo, CA: Morgan Kaufmann.

SHOUP, D. (2014) The High Cost of Minimum Parking Requirements, in: ISON, S.;

SIRMANS, G. S.; MACDONALD, L.; MACPHERSON, D. A.; ZIETZ, E. N. (2006) The Value of Housing Characteristics: A Meta Analysis. **The Journal of Real Estate Finance and Economics**, v. 33, n. 3, p. 215-240.

WELCH, T. F.; GEHRKE, S. R.; WANG, F. (2016) Long-term Impact of Network Access to Bike Facilities and Public Transit Stations on Housing Sales Prices in Portland, Oregon. **Journal of Transport Geography**, v. 54, p. 264-272.

WITTEN, I. H.; FRANK, E. (2005) **Data Mining: Practical Machine Learning Tools and Techniques**, 5 ed. Boston, MA: Morgan Kaufmann.

WOODS, E.; KYRAL, E. (1997) **Ovum Evaluates Data Mining**, London: Ovum.

XIAO, Y.; ORFORD, S.; WEBSTER, C. J. (2016) Urban Configuration, Accessibility, and Property Prices: A Case Study of Cardiff, Wales. **Environment and Planning B: Planning and Design**, v. 43, n. 1, p. 108-129.